Mozambique - Farmer Income Support

Report generated on: January 7, 2016

Visit our data catalog at: https://data.mcc.gov/evaluations/index.php

Overview

Identification

COUNTRY
Mozambique

EVALUATION TITLE
Farmer Income Support

EVALUATION TYPE
Independent Impact Evaluation, Independent Performance Evaluation

ID NUMBER
DDI-MCC-MOZ-ABT-FISP-2014-v2

Version

VERSION DESCRIPTION
Anonymized dataset for public distribution

Overview

ABSTRACT

Trees

For the epidemic zone, the evaluation estimated the impact of FISP on disease prevalence and estimated the consequent impact on coconut production and farmer incomes. In addition, the evaluation also assessed the pre-conditions for future increase in income by examining whether the seedlings planted through FISP survived.

To assess these impacts, because the ideal approach of randomly assigning treatment to farmers was not feasible, the evaluation employed the next best approach. It identified geographic areas that were very similar at baseline and were otherwise split into treatment and comparison areas by some external “quasi-random” factor (the phytosanitary barrier), and matched a subset of these areas on disease prevalence and distance to coast. This approach is sometimes called a “border discontinuity” design in the applied microeconomics literature. As mentioned in Section 2.3, the phytosanitary barrier was a north-to-south imaginary line that FISP delineated, separating the areas with less than 10 percent disease prevalence (on the west side of the barrier) from the areas with more than 10 percent disease prevalence (on the east side). FISP worked on the west side of the barrier. Defined this way, the project worked in the entire epidemic zone.

However, in effect this barrier was imperfect because while it is true that generally (or on average) the disease prevalence is greater near the coast, the disease prevalence has a mosaic pattern. This means that even on the east side of the barrier, there can be areas with disease prevalence lower than 10 percent. Furthermore, the barrier was imperfectly drawn even along the perceived divide of 10 percent prevalence, as it was drawn along natural barriers such as water bodies and roads. These characteristics of the barrier and how the FISP treatment areas were defined give a unique opportunity to identify the counterfactual areas in the areas just east of the barrier. Accordingly, we used the 2008 TTI aerial photography data (TTI Production, 2009) to identify areas with similar disease prevalence as proximate as possible to either side of the barrier, with the areas on the west side serving as treatment areas, and areas on the east side serving as the comparison areas .

From among all FISP project areas on the east side of the barrier, and the non-project areas on the west side of the barrier, we identified areas where the disease prevalence was between 0 and 15 percent. These treatment and comparison areas amounted to 93 project and 65 non-project census enumeration areas (EAs) very close (within 73 kilometers) to the phytosanitary barrier that shared similar disease prevalence. Within these areas, we further refined the sample and identified approximately 24 nearest neighbor pairs of EAs matched on baseline disease prevalence and distance to coast. Our sample selection procedure, which resulted in 48 EAs, limits analysis and direct applicability to areas close to the phytosanitary barrier. The pairs were matched on:

Same district

Average baseline CLYD prevalence in 2008, in the 5-kilometer radius

Distance to coast.

No counterfactual EA was “matched” with more than one treatment EA. A small wrinkle in this identification strategy was that the FISP phytosanitary barrier was reevaluated a total of five times throughout the project, causing the barrier to shift during the course of the implementation (see Figure 2-1). To ensure that all comparison areas did not receive treatment, we have chosen comparison located east of the 2012 barrier, which is the furthest east the barrier was shifted over the course of the project.

It is important to note that in three out of the eight FISP districts there was no barrier drawn: in Pebane, Moma, and Angoche, the barrier was the coast, and FISP worked in these districts in a way that made it difficult to identify non-project areas. There was no barrier drawn in the district of Chinde, but epidemic and endemic zones (as well as project and non-project areas) were geographically separated in a way that allowed us to include the district in our sampling frame. As a result, our quantitative evaluation is focused on five out of the eight FISP districts (Nicoadala, Namacurra, Maganja da Costa, Chinde, and Inhassunge). In the remaining three districts in the Nampula district, we relied on a qualitative survey to draw lessons about the project's efficacy.

In sum, the phytosanitary barrier distinguished between epidemic FISP areas and epidemic non-FISP areas. We identified 93 FISP census EAs on the west side of the barrier and 65 non-project EAs on the east side of the phytosanitary barrier but very close (within 73 kilometers) with similar baseline disease prevalence (using the 2008 aerial photography data from TTI Production, 2009). From these, 24 pair-wised matched project and non-project areas serve as our treatment and comparison groups. Their baseline similarity, geographical similarity, and haphazard separation by the phytosanitary barrier create the ideal conditions for a quasi-experimental evaluation. This quasi-random approach identifies the impact because difference in outcomes can be attributed only to the intervention; other potentially confounding factors are similar across the phytosanitary barrier, due to geographic proximity and because of matching (Section 5.3 provides more details on the results of the matching).

Households

We designed two separate quantitative evaluations to assess the activities in the epidemic zones and the endemic zones.

Evaluation of Epidemic Zone Program: In the epidemic zones, the evaluation used a border discontinuity design. Specifically, it identified geographic areas that were very similar at baseline, except that they were split into treatment and comparison areas by an external “quasi-random” factor-the phytosanitary barrier. The phytosanitary barrier was a north-to-south imaginary line that FISP delineated, separating the areas with less than 10 percent disease prevalence (on the west side of the barrier) from the areas with more than 10 percent disease prevalence (on the east side). FISP did not conduct any epidemic zone activities east of this barrier. Despite the clear distinction of the barrier for implementation activities, the disease had a mosaic pattern, implying that the barrier was imperfect at delineating CLYD prevalence levels. This feature aided our evaluation, because we were able to select comparison areas to the east of the barrier that had roughly the same qualifications for selection (baseline disease prevalence) as the treatment areas. Among a subset of the census enumeration areas (EAs) on both sides of the barrier that had less than 15 percent disease prevalence, we matched EAs based on distance from coast (which also proxies for distance to barrier) and baseline disease prevalence. This strategy was feasible in five out of the eight FISP districts. In the remaining three districts in the Nampula province, where the epidemic and endemic zone programs were combined and there was no barrier separating project areas, we relied on qualitative data to assess the project.

To assess impact on outcomes related to household income coconut production and seedling survival, the unit of analysis was households. We sampled 800 households across 24 treatment and 24 comparison EAs. After accounting for nonresponse, missing observations, and outliers, we obtained a study sample of 666 usable observations. For each household we gathered information on their pre-intervention demographic characteristics and information to measure the key outcomes: seedling survival, coconut production, and income.

To assess the impact of FISP on disease prevalence, we surveyed 16,000 trees that were sampled from the same enumeration areas as the household survey in the epidemic zones. We first identified the land area within the matched EAs that had tree cover, based on the 2008 aerial photography data. Within these areas, we randomly sampled 400 grids, each 100 by 100 meters, and within each grid we drew a sample of 40 trees in four clusters. For each tree in the grid, the tree enumerators visually inspected whether the tree was healthy; had only CLYD (and not beetle infestation); had beetle infestation only; or had CLYD and beetle infestation. The enumerators assessed beetle infestation also because CLYD is closely associated with increase in beetle infestation since dead trees - dead because of CLYD - can become a breeding ground for beetles.

To answer the question on FISP's impact on disease spread, however, we could not apply this method because that requires time series information on disease spread before, during, and after FISP. We conducted the analysis on two pairs of treatment and comparison areas for which we could find satellite data to estimate the prevalence of healthy trees between 2008 and 2014.

Evaluation of the Endemic Zone Program: In the endemic zones, which were entirely east of the phytosanitary barrier, the evaluation used a non-experimental method for selecting untreated geographic areas that, at baseline, were very similar to the treated areas. We matched pairs of treatment and comparison enumeration areas on baseline disease prevalence and distance from coast. The results of our matching and comparison of their self-reported pre-intervention characteristics suggest that the two groups are comparable and did not require any re-weighting to “look similar.”

In the endemic zone, we surveyed 720 households in 80 EAs across all districts. Of the 720 households surveyed, 561 observations were usable. In addition to information gathered for the epidemic zone farmers, for endemic zone farmers we also gathered information on their adoption, production, and sale of crops promoted by FISP.

Other Questions: To answer other questions, including the assessment of the BDF activity, the R&D activity, and the adaptation and adherence of FISP implementation to project goals, we conducted a desktop review of the literature. In addition, for the BDF activities and assessment of FISP's adaptation and adherence to project goals, we conducted interviews with key stakeholders, and focus group discussions and interviews with farmers and grant recipients.

EVALUATION METHODOLOGY
Ex-Post with comparison group, Independent Ex-post ERR, Performance Evaluation

UNITS OF ANALYSIS
Households, Clusters of Coconut Palms

KIND OF DATA
Sample survey data [ssd], Administrative records data [adm], other

TOPICS

Topic	Vocabulary	URI
Agriculture and Irrigation	MCC Sector	
Gender		

KEYWORDS
Impact Evaluation, Coconut Lethal Yellowing Disease, Farmer Income Support Project, Mozambique, Performance Evaluation, Case study, Economic Rate of Return

Coverage

GEOGRAPHIC COVERAGE
Zambezia and Nampula Provinces; Chinde, Nicoadala, Namacurra, Maganja da costa and Inhassunge.

Rural areas only.

UNIVERSE
Rural households living in the "coconut belt" of central Mozambique, specifically in the coastal districts of Zambezia and Nampula provinces.

For the biophysical survey, the study population are coconut palms in the same region.

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Abt Associates	

FUNDING

Name	Abbreviation	Role
Millennium Challenge Corporation	MCC	Review of Metadata

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Abt Associates	Abt		Creation of Metadata

DATE OF METADATA PRODUCTION
2015-05-08

DDI DOCUMENT VERSION
Version 2, May 2015.

DDI DOCUMENT ID
DDI-MCC-MOZ-ABT-FISP-2014-v2

MCC Compact and Program

COMPACT OR THRESHOLD
The Compact’s goal was to reduce poverty in Mozambique through economic growth and increase economic opportunities for Mozambicans living in the northern region. The Program Objective was to increase the productive capacity of the population in selected provinces in northern Mozambique, with the intended impact of reducing the poverty rate, increasing household income, and reducing chronic malnutrition in the targeted districts.

PROGRAM
The objectives of FISP were to improve the productivity of coconut products and encourage diversification into production of other cash crops. The project was designed to eliminate biological and technical barriers hindering economic growth among farms and targeted enterprises located in the Compact area's eastern coastal belt (Zambézia and Nampula provinces). It aimed to reduce the prevalence and economic impact of CLYD by supporting 1) the cutting and burning of infected trees; 2) the planting of new, less-susceptible seedlings; 3) community education and awareness programs to assist coconut growers in preventing the disease in the future; and 4) crop diversification technical assistance.

MCC SECTOR
Agriculture and Irrigation (Ag & Irr)

PROGRAM LOGIC
The Compact's goal was to reduce poverty in Mozambique through economic growth and increase economic opportunities for Mozambicans living in the northern region. FISP's overall objective was to increase the agricultural capacity of the population in selected provinces in northern Mozambique, with the intended impact of reducing the poverty rate, increasing household income, and reducing chronic malnutrition in the targeted districts. Specifically, the project was designed to eliminate biological and technical barriers hindering economic growth among farms and targeted enterprises located in the Compact area's eastern coastal belt of Zambézia and Nampula provinces, with greater emphasis of the program in Zambezia. The FISP activities were designed to reduce the adverse impact of CLYD on farmer income. Also, in epidemic zones FISP sought to reduce the loss of coconut production, while in endemic zones it sought to encourage diversification into production of other cash crops. It aimed to reduce the economic impact of CLYD through the five activities.

PROGRAM PARTICIPANTS
FISP is expected to have benefitted 277,763 smallholders in the coconut belts of Zambézia and Nampula by 2028. Of these, it is estimated that 119,273 individuals are in the endemic areas and 158,390 are in the epidemic areas. Among the beneficiaries, FISP provided targeted technical assistance to over 3,000 smallholders to mitigate significant income loss due to CLYD and to assist in improving the quality of other crops planted on their holdings. According to the program logic, the estates did not get direct technical assistance they were expected to benefit from the disease eradication.

Sampling

Study Population

Rural households living in the "coconut belt" of central Mozambique, specifically in the coastal districts of Zambezia and Nampula provinces. For the biophysical survey, the study population are coconut palms in the same region.

Sampling Procedure

Household Survey:

The unit of analysis for the income impact is households living within the selected enumeration areas. The study sample for the analysis to assess FISP's impact on coconut production, farmer income, and seedling survival in the epidemic zones comes from a survey of 666 households. We drew the sample of households by first selecting households within villages from EAs that were subject to the FISP intervention. As described above, we refined our sample of EAs from 93 project and 65 non-project census EAs very close (within 73 kilometers) to the phytosanitary barrier that shared similar disease prevalence, by finding approximately 10 nearest neighbor pairs of EAs matched on baseline disease prevalence and distance to coast in each district. Within each selected EA, we conducted a random walk to select eight farmers to interview. Based on our power analysis calculations, we planned to survey 800 households; due to accessibility of households or consent to interview we ended up with a sample of 728 households. Of the 728 households surveyed, 666 observations were usable after accounting for missing observations or outliers. Respondents answered questions about their demographics, education, household structure type and ownership of various durable commodities (e.g., mobile phone, bicycle) as measures of wealth, agricultural production in 2009 prior to the implementation of FISP, agricultural output and sales from the most recent 12 months (August 2013-July 2014), household members' other income-generating activity in the most recent 12 months, and their knowledge of CLYD.
In addition to the quantitative data that covered all but three FISP districts (Pebane, Moma, and Angoche), we also collected data using qualitative methods for all targeted districts with a special focus on the three districts that were not included in the quantitative analysis. Overall, we conducted 21 focus group discussions and 13 interviews (4 in comparison areas) with smallholder farmers in the epidemic zones with the intent to explore in-depth how the farmers in FISP targeted areas perceived each component of the project, what changes they saw in their lives over the course of the project and specifically on disease progression, coconut production, and income. The qualitative data helped us enrich the results of the quantitative data and provided us with details on Nampula.
As for the evaluation in the endemic zones, the unit of analysis for the income impact is the people living within EAs, based on the population of people who live in those areas. Informed by our power analysis calculations during the design phase, we surveyed 720 households in 80 EAs across all districts. Of the 720 households surveyed, 561 observations were usable.
In the endemic zones, we conducted seven focus group discussions and seven interviews with smallholder farmers to explore in-depth how the farmers in FISP project areas perceived each component of the project, what changes they saw in their lives over the course of the project and specifically changes in disease progression, coconut production, and income. The qualitative data helped us enrich the results of the quantitative data.

Biophysical Survey:

Our study sample for the impact analysis on disease prevalence draws from the same treatment and comparison areas as that for the income analysis in the epidemic zones-the areas on either side of the phytosanitary barrier. To study FISP's impact on CLYD, and to determine the ideal unit of analysis, an understanding of how the disease spreads across trees (the disease vectors) is needed. Unfortunately, little is known about the progression of CLYD among trees and about the possible disease vector; hence, it is difficult to know what the ideal unit of analysis should be. It is conceivable that "tree clusters"-trees separated by physical barriers such as land without trees or water bodies-might be a reasonable way to delineate units of analysis. Even if this approach had sufficient scientific basis, it would require a full listing of all clusters that exist in the FISP area- a task that is infeasible given the labor effort required.
Without a good scientific basis for using tree cluster as the unit of analysis, and knowing that trees within villages will be related at least in how they are cared for, we consider it more reasonable to use the same unit of analysis that we use for the income impact. Instead of using EAs, however, we use a level higher assuming that trees are connected to each other over larger areas. Therefore, we use villages (nomes) as our units of analysis, and propose to sample one-hectare square grids from the villages. We will conduct this analysis in a sub-sample that is selected for the income analysis and the strategy therein to select matching treatment and comparison groups. Within each selected village, we will enumerate one-hectare square grids that have a specified minimum density of coconut trees (using baseline aerial photography), and we will randomly select the required number of grids from this list of sample-eligible grids. We propose to conduct biophysical measurement of the disease in 800 to 1,600 units, equally divided across treatment and comparison areas. For efficiency, we will conduct this analysis in a sub-sample of the study sample for the income analysis. The final decision on the total number of observations will depend on the cost of survey and the amount we need to have available for the case

study to assess the impact of FISP on disease spread rate. The intervention would have to have an average impact between 15.1 and 15.3 percent in the epidemic region in order to have an 80-percent chance of being detected in our analysis. The monitoring data suggest that the lack of any tree-cutting activity can result in an impact of more than 19 percent; thus, the MDI seems appropriate. If we see an impact well above 19 percent, it will be evidence of FISP's impact in curtailing the disease spread rate.

Deviations from Sample Design

Household Survey

The only deviation from the original sampling plan for the household survey was the number of EAs used for data collection. After discussion with the survey firm, we determined that it was more cost effective to reduce the number of EAs and increase the number of households surveyed in each EA. As a result, the survey firm was tasked to complete interviews with 9 households in 80 EAs for the endemic zone and 9 households in 100 EAs for the epidemic region. The total sample size was not affected greatly moving from the original 1536 to 1520. Consequently, the MDI for the endemic region shifted from 2,423 meticales (USD75)) to 2,634 meticales (USD82). In the Epidemic zone it changed from 3,857 meticales (USD120) to 4,056 (USD 127).

Biophysical Survey:

The sample to study the impact of FISP on disease prevalence comes from a survey of 16,000 trees that were sampled from the same treatment and comparison enumeration areas as the household survey in the epidemic zones. We first identified the areas within the matched EAs that had at least 50 percent tree cover, and baseline disease prevalence less than 10 percent using the 2008 aerial photography data. Within these areas, we sampled 400 grids each 100 by 100 meters, equally divided between treatment and comparison areas. Within each grid we drew a sample of 40 trees in four clusters as follows: Locate the tree nearest the Global Positioning System (GPS) coordinates of the center of the sample grid, which served as the primary reference palm tree, and identify nine additional trees within a radius of approximately 30 meters.
Walk approximately 50 m north to a new area with palm trees and repeat as above by selecting a secondary reference palm tree and record data. Return to the primary reference palm tree, walk 50 m southeast to a new area with palm trees, and repeat as above by selecting a next secondary reference palm tree and record data. Return to the initial primary reference palm tree, walk 50 m southwest to a new area with palm trees, and repeat as above by selecting another secondary reference palm tree and record data.
For each tree in the grid, the tree enumerators visually inspected whether the tree was healthy; had only CLYD (and not beetle infestation); had beetle infestation only; and whether the tree had CLYD and beetle. If the tree had beetle infestation, the tree enumerators gave an ordinal indication of high, medium, or low infestation. The enumerators also visually assessed the height of the tree and number of coconuts that the trees had as an ordinal measure-zero, high, medium, or low number of coconuts. Appendix D provides the complete survey sheet.
Because the program implementers and evaluation relied on visual identification of CLYD, as part of the evaluation we also assessed whether visual identification is a close approximation of actual CLYD prevalence (i.e. that visual identification does not yield a high portion of false positives and false negatives). The program itself had evidence indicating the reliability of a positive identification of CLYD based on visual symptoms, which are backed up by the molecular studies of the FISP research component. However, the project made no robust assessment to validate a false negative rate for visual identification (i.e. the tree looks CLYD-free, but is actually asymptomatically infected by the causal agent of CLYD). Thus, FISP did not provide any way to assess the proportion of visually healthy trees infected with the causal agent of CLYD.
Our analysis finds that a visual identification of CLYD has a 92.4 to 99.8 percent probability of being accurate. However, our analysis also finds that of the trees that are observed as healthy (i.e., without visible CLYD symptoms), 1.4 to 7.6 percent of these can be expected to be infected with the causal agent of CLYD. We have no knowledge of the contribution of disease spread between trees that are observed as healthy but infected with the causal agent of CLYD vis-à-vis trees that have visible CLYD symptoms.

Response Rate

Household Survey

The power analysis assumed a response rate of 90%. Our final calculated sample size was 1520. By the end of the survey, we were provided data on 1513 households. The evaluation team then dropped some observations while cleaning, bringing the final sample size to 1343, which is approximately 89% of the original sample size. Our power analysis was constructed to

allow a 10% reduction of sample size without affecting the MDIs.
Based on the assumptions of our original power analysis, this sample size decrease results in less than a half of a percent increase in MDI. Moreover, our power analysis assumptions were based on limited available data. We will report the actual MDEs concurrently with our findings report; the actual MDEs will be calculating using the standard error of the impact estimates.

Biophysical Survey

Since trees are the unit of measure, there were not response rate concerns

Weighting

No survey wieghts have been applied to this data.

Questionnaires

Overview

Household questionnaire:
The household questionnaire was developed based on the national agricultural survey (TIA) with some modifications and additions to focus the content on coconut production and FISP activities.

Qualitative questionnaires:
Rough drafts of the qualitative questionnaires were developed by the evaluation team to highlight the main pieces of information needed from stakeholders. The qualitative survey firm then used that information to create final versions of "interviewer guides" used to facilitate discussion in one-on-one interviews, focus group discussions and value chain assessments.

Biophysical data collection instrument:
This was not a questionnaire, but a data collection instrument used to count trees. Enumerators (separate from those who conducted the household survey) filled these out as part of the biophysical survey in order to establish an endline disease rate in the FISP intervention areas.

All questionnaires are provided as external resources.

Data Collection

Data Collection Dates

Start	End	Cycle
2014-08-18	2014-10-30	Household Survey
2014-08-18	2014-09-10	Biophysical Survey
2014-09-01	2014-11-27	Qualitative Survey

Data Collection Notes

Household Survey:

The pretest for the survey took place on and included 42 enumerators who were participating in the training. This group of enumerators was later funneled down to 36, based on performance during the pretest and scores on a written final exam. Enumerators conducted the pretest individually, with a supervisor assisting at the beginning of each interview by making introductions and ensuring the respondent was aware and willing to participate.

Interviews averaged approximately 45 minutes and were conducted primarily in Portuguese and Chuabo. Two Abt evaluation team members and 4 survey firm team members also participated by accompanying enumerators during the interviews and coordinating with village authorities to get permission to interview.

Biophysical Survey:

The pretest for the survey took place from August 7-9, 2014 and included the two survey managers as well as the "enumerators" or tree counters and one member form the Abt evaluation team. During the pretest, the survey team tested the counting methodology, GPS devices and data entry forms, ensuring each tree counter understood the process and entered the data correctly. At the end of the pretest, all tree counters demonstrated satisfactory performance and the methodology was confirmed to be effective.

Questionnaires

Household questionnaire:
The household questionnaire was developed based on the national agricultural survey (TIA) with some modifications and additions to focus the content on coconut production and FISP activities.

Qualitative questionnaires:
Rough drafts of the qualitative questionnaires were developed by the evaluation team to highlight the main pieces of information needed from stakeholders. The qualitative survey firm then used that information to create final versions of "interviewer guides" used to facilitate discussion in one-on-one interviews, focus group discussions and value chain assessments.

Biophysical data collection instrument:
This was not a questionnaire, but a data collection instrument used to count trees. Enumerators (separate from those who conducted the household survey) filled these out as part of the biophysical survey in order to establish an endline disease rate in the FISP intervention areas.

All questionnaires are provided as external resources.

Data Collectors

Name	Abbreviation	Affiliation
Polaris Ltd.		Survey Firm
Josef Pudivitr		Consultant

Supervision

Household Survey:

Rafael Achila--Survey Field Coordinator/Supervisor; oversaw all field work and ensured logistics for teams were taken care of.
Arlindo Miguel--Survey Field Coordinator/Supervisor; oversaw all field work and ensured logistics for teams were taken care of.
Luis Lopes-- Data entry coordinator; responsible for managing the data entry personnel and deliver the data in its final format.
Anabela Mobata--Survey Manager; main point of contact and responsible for all survey activities.

Interviewers were conducted by individual enumerators, who worked in teams of 5; 4 enumerators and one controller who was meant to directly supervise the enumerators and ensure data was entered correctly. The survey firm was managed by Mikal Davis, from Abt Associates.

Biophysical Survey:

Josef Pudivitr--Survey Manager; Overall survey implementation lead.
Julio Kayibanda--Survey supervisor; managed tree counters and coordinated logistics.

Six tree counters were used for the survey and 3 technicians used for tree sample extraction. All tree counters were managed by Mr. Kayibanda and the tree extraction team were a group of experts from the national institute for agricultural research. This team was managed directly by Dr. Julian Smith, from Food and Environment Research Agency (FERA).The biophysical survey was managed by Dr. Julian Smith.

Data Processing

Data Editing

For both surveys:
Data editing took place at a number of stages throughout the processing, including:
a) Field collection
b) During data entry
c) Structure/completeness checking
d) Secondary editing
e) Analysis

All data cleaning takes place in “CreatePublicUseDatasets.do” (which takes the raw data as input) and “CreateAnalysisDataset.do” (which takes the public use data as input). Some cleaning does not occur until the Analysis Dataset is made: a lot of analysis is required in order to determine total land area, sales per FISP crop, etc. We removed some outliers in the “CreateAnalysisDataset.do” file; this file also saves a file with all of the household ids that remain after the final cleaning. We then circle back to the “CreatePublicUseDatasets.do” file, and make sure that the public use data sets only include the data that survived the final cleaning in the “CreateAnalysisDataset.do” file.

Other Processing

Household Survey:

Data was entered over a period of 12 days using CSpro. 5 data entry staff imputed the data. The data entry staff averaged approximately 25 surveys a day using individual laptops in the field. 100% verification of all variables was performed using independent verification, i.e. double entry of data.

Biophysical Survey:

Data was entered over a period of one week by the survey manager, Josef Pudivitr. No double entry occurred, but quality control was performed by the Abt team using the hard copies of the data collection sheets. A subset of sheets were randomly chosen and compared to the dataset submitted by Mr. Pudivitr to confirm the quality was acceptable.

Data Appraisal

Estimates of Sampling Error

Will be provided after analysis is complete.